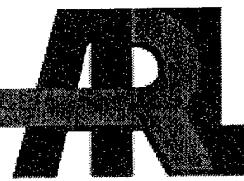ARMY RESEARCH LABORATORY

Glossary and Catalog of MeT Data Representation

by Edward Measure

ARL-TR-2712

February 2003

20030320 052

Contents

1

Preface

This report is intended to catalog the major digital formats in wide use for representing meteorological data. Each data format discussed is briefly defined, and references are given to provide a more detailed discussion for each format. The author intends to follow this report with another report that will discuss, in considerably greater detail, the most important formats used for satellite data. No claim is made that the list of formats given in this report is exhaustive, and the author would appreciate readers calling attention to important formats that have been neglected or overlooked for incorporation into future editions. The author can be contacted at: *emeasure@arl.army.mil*

This report is most usefully viewed as an electronic hypertext document, since, wherever possible, the references given are available on the Internet, and the appropriate Uniform Resource Locators (URLs) are given.

Executive Summary

A vast amount of data is required for the description of the atmosphere to represent its complexities and the range of scales over which important phenomena occur.. Numerical models, satellites, and other measurements generate very large quantities of data. A wide variety of specialized computer formats has been used for storage and interpretation of that data. This report contains brief descriptions and references for further information of the primary data representation schemes.

1. Introduction

The atmosphere is one of the most complex physical systems studied and phenomena of interest occur over a vast range of scales. Attempts to understand the atmosphere and its behavior have led to collection of a rapidly increasing volume of measured data from diverse types of sensors. Making use of this data requires systematic methods for storing, organizing, analyzing, and comparing these various measurements. An important aspect of the management of this data is the system used to record and organize our data, that is, the data representation scheme.

The proliferation of measurements, and the availability of digital media to store and transmit the volumes of data generated by them, has led to the development of a wide variety of formats and data representation schemes. Many of these formats and schemes are specialized or restricted to a single operating system, programming language, or machine type. Several attempts have been made to bring some order and standardization to the resulting "Tower of Babel" of data formats in the form of "standard" data formats.

2. Standard Formats

This report presents a glossary and catalog of popular and standard formats used in meteorology and related sciences, with special emphasis on the large-scale digital data sets that are now the mainstream of our science. Each format discussed in this report is accompanied by a brief description and one or more references to more complete descriptions. In most cases, the referenced material is available online; the appropriate URL addresses are included with the descriptions.

AVHRR (Advanced Very High-Resolution Radiometer Level 1b data format.) Kidwell, (1999) states, "Level 1b is raw data that have been quality-controlled, assembled into discrete data sets, and to which Earth's location and calibration information have been appended (but not applied)." AVHRR data levels do not align precisely with Earth Observing System (EOS) levels. (See below).

CANDIS. An antecedent format of netCDF, developed by Dave Raymond of the New Mexico Institute of Mining and Technology (Raymond, 1988).

BUFR (Binary Universal Form for the Representation.) "BUFR is an acronym for Binary Universal Form for the Representation of meteorological data. BUFR is a World Meteorological Organization (WMO) standard binary code for the exchange and storage of data. The format is documented in WMO Manual on Codes; WMO Publication No. 306; Volume I, Part B; 1995 Edition, plus Supplement 1." Like most of the other modern digital formats, BUFR aims to be self-describing, that is, the data files contain information describing the nature of their contents. A good source of information on the use of BUFR is W. Thorpe's A Guide to the WMO Code Form FM 94 BUFR.

(Thorpe, undated). Stackpole (1993) presents a wider ranging discussion of BUFR and its sister standard gridded binary (GRIB) (see below), including some notes on history and philosophy.

CDF Common Data Format. CDF is a data interface, format, and associated software developed at the National Aeronautics & Space Administration (NASA) Goddard National Space Science Data Center. "The National Space Science Data Center's (NSSDC) Common Data Format (CDF) is a self-describing data abstraction for the storage and manipulation of multidimensional data in a discipline-independent fashion ... CDF has its own internal self describing format, it consists of more than just a data format. CDF is a scientific data management package (known as the CDF Library) which allows programmers and application developers to manage and manipulate scalar, vector, and multidimensional data arrays." (King, 2001)

EOS Data Product Levels. EOS data levels do not align precisely with AVHRR data levels. "Data levels 1 through 4 as designated in the Product Type and Processing Level Definitions document. . .

- Raw Data - Data in their original packets, as received from the observer, unprocessed by EDOS.

- Level 0 - Raw instrument data at original resolution, time ordered, with duplicate packets removed.

- Level 1A - Reconstructed unprocessed instrument data at full resolution, time referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters (i.e., platform ephemeris) computed and appended, but not applied to Level 0 data.

- Level 1B - Radiometrically corrected and geolocated Level 1A data that have been processed to sensor units.

- Level 2 - Derived geophysical parameters at the same resolution and location as the Level 1 data.

- Level 3 - Geophysical parameters that have been spatially (sic) and/or temporally re-sampled (i.e., derived from Level 1 or Level 2 data).

- Level 4 - Model output and/or results of lower level data that are not directly derived by the instruments." (Ullman, 2000)

DPEAS (Data Processing and Error Analysis System). DPEAS is a proprietary system created by Andrew S. Jones of Colorado State University. "DPEAS was created to overcome the inherent difficulties of working with multiple data formats." (Jones, 2001). The system converts many types of data to hierarchical data format (HDF)-EOS format and has other data-processing and error-analysis capabilities.

FAN (File Array Notation). FAN high-level interface to NetCDF data. FAN (File Array Notation) is an array-oriented language for identifying data items in files for the purpose of extraction or

modification." (Davies, 1996). Apparently, FAN only supports NetCDF formats but in principle it should be extensible to other formats.

GRIB (GRIdded Binary). "GRIB and BUFR were created by members of the international meteorological community, with some oceanographers in tow, in an effort to standardize the exchange of their data, in both "processed" forms (gridpoint analysis and forecast information, generally) and observational forms, using modern, bit transparent, communications protocols. Both formats have been accepted by the WMO, which retains a standing committee that oversees changes, augmentations, and improvements to the code forms. The WMO "acceptance" raises the code forms to the level of an international standard, at least within the meteorological community." (Stackpole, 1993). GRIB is documented in WMO publication #306, Manual of Codes. (American Meteorological Society, 1998) "Both GRIB and BUFR share a common "model": each GRIB or BUFR record (or message, if you are considering their use in a communication context) contains all the information needed to properly decode the data, recognize the nature of the information, and place it in its proper time and space location(s). This is accomplished generally by the inclusion of specific numeric values, or by reference to external tables that define the "meaning" of the parameters included in the record. The WMO looks after maintaining and updating those tables." (Stackpole, 1993) The European Centre for Medium-Range Weather Forecasts (2001) provides links to some software for encoding, decoding, manipulating, plotting, and converting GRIB data.

HDF. HDF is a format that was developed by the National Center for Supercomputing Applications (NCSA) and is used in environmental science and also in " . . . Neutron Scattering, Non-Destructive Testing, and Aerospace, to name a few." (NCSA, 2001b). HDF is intended for very large data sets and has been adopted for NASA's Earth Observing System satellites that will generate data rates in excess of a terabyte per day. HDF exists in two incompatible forms, HDF4 and HDF5. The HDF5 form is the newer and more capable version. An online reference manual for HDF4 (in .pdf or post-script format) is available at NCSA (1995). An HDF5 tutorial is online at NCSA (2001a); access to a vast amount of additional documentation is available through the links at NCSA (2001b). HDF is available for Windows, LINUX, and many other operating systems and has prominent users who include NASA and the U. S. Army Research Laboratory.

HRPT (High Resolution Picture Transmission). "The Advanced Very High Resolution Radiometer (AVHRR) is a broad-band, four or five channel (depending on the model) scanner, sensing in the visible, near-infrared, and thermal infrared portions of the electromagnetic spectrum. This sensor is carried on NOAA's Polar Orbiting Environmental Satellites (POES), beginning with TIROS-N in 1978." (Advanced Very High Resolution Radiometer)

"HRPT data are full resolution image data transmitted to a ground station as they are collected." (Advanced Very High Resolution Radiometer)

McIDAS (Man-computer Interactive Data Access System). McIDAS is a proprietary software and hardware system focused on the ingestion, display, and storage of environmental data, especially satellite data. According to the description at the University of Wisconsin Space Sciences Engineering

Center web site (http://www.ssec.wisc.edu/software/mcidas.html) "McIDAS (Man computer Interactive Data Access System), under development since 1970 at the University of Wisconsin-Madison's SSEC, is a sophisticated, video interactive set of tools for acquiring, managing, analyzing, displaying and integrating environmental data." The URL referenced above includes links to more detailed descriptions of McIDAS, its components, and functional capabilities.

METAR/TAF – METAR/SPECI. METAR/TAF – METAR/SPECI is "... the international standard code for hourly and special surface weather observations..." according to Jarvi (1996). METAR is a highly telegraphic string of letters and numbers that aims to distill considerable information into a compact but still (barely) human readable format. A sample observation is given as follows: IAD SA 1055 11 SCT E15 OVC 1/2S-F 045/33/29/2119G27/945/R04VR30 PK WND 1929/16. TAF is a related standard for airport forecasts. Although the standard is international, U.S. usage is exceptional in using some nonstandard units such as feet and knots instead of the SI standards.

The exact derivation of the acronym METAR is unclear, with Jarvi stating rather vaguely "The METAR acronym roughly translates from the French as Aviation Routine Weather Report." On the other hand, a French source claims (rather plausibly) that "M.E.T.A.R. sont les initiales d'une expression anglaise: METeorological Airport Report," (MétéoSum, 2001). (M.E.T.A.R are the initials of an English expression: METeorological Airport Report). Jarvi, cited above, gives a detailed explanation of the format.

NetCDF (Network Common Data Format). "NetCDF (network Common Data Form) is an interface for array-oriented data access and a library that provides an implementation of the interface. The netCDF library also defines a machine-independent format for representing scientific data. Together, the interface, library, and format support the creation, access, and sharing of scientific data. The netCDF software was developed at the Unidata Program Center in Boulder, Colorado" (Rew, 2001a). The NetCDF format aims to be self-describing, and machine independent and is "... is an interface for array-oriented data access and a freely distributed collection of software libraries for C, Fortran, C++, Java, and perl" (Rew, 2001b). Access to freely available software and extensive documentation is available though links at Rew (2001b) cited above.

NOAAPORT. NOAAPORT is a satellite-based broadcast system for dissemination of NOAA environmental data, including satellite and other data. "The NOAAPORT broadcast system provides a one-way broadcast communication of NOAA environmental data and information in near-real time to NOAA and external users. This broadcast service is implemented by a commercial provider of satellite communications utilizing C-band." Data from various sources is up-linked to satellite and then broadcast down to users with suitable receivers and appropriate software. (Jarvi, 2001a) . (http://205.156.54.206/noaaport/html/overview.shtml) "Weather data is collected by GOES satellite environmental sensors and NWS observing systems, and processed to create products. The products are fed to the AWIPS Network Control Facility (NCF) which routes the products to the appropriate NOAAPORT channel for uplink and broadcast." (op. cit)

NOAAPORT data streams are mostly in WMO format. A more detailed description of the associated data formats is discussed by Jarvi (2001b)., and references therein.

SOAP (Simple Object Access Protocol). SOAP is documented by Gudgin, et. al. (2001).

TDF (TeraScan Data Format). TDF is a proprietary data format developed by Seaspace Corporation. This data format is used by Seaspace's Terascan software for storage and manipulation of a variety of environmental data, including polar orbiter and geostationary satellite data. The TeraScan format was in many respects the model for NetCDF. The following description is given by Seaspace (1999): "The TeraScan software consists of several hundred TeraScan functions—UNIX commands developed especially for capturing data from remote-sensing environmental satellites and processing, distributing, and displaying the data on the TeraScan system." The Terascan TDF data format is documented on the web site (Seaspace, 2000); other document links are also available.

XDR (External Data Representation) Standard. XDR is an Internet Draft Standard Protocol (Reynolds, et. al., 2001), for representing digital data in a machine-independent form. XDR is used by netCDF (Rew et. al., 1996), and other data formats for machine-independent data representation. XDR is described in detail in RFC 1832 (Srinivasan, 1995).

Upper-Air Data Codes. Standard codes for radiosonde (FMH#4) and upper wind data (FMH#6) are described in detail in the referenced Federal Meteorological Handbooks and the World Meteorological Publication (AMS, 1996). These codes, like the METAR format, are a compromise between concision and human readability dating back to a pre-computer age, a telegraphic mixture of characters and numbers that is completely obscure to the untrained eye, and, in my opinion, an unfortunate fossil that should be consigned to the dustbin of history as soon as practicable.

XML (Extensible Markup Language). XML according to the official XML web site, http://www.w3.org/XML/, "The Extensible Markup Language (XML) is the universal format for structured documents and data on the Web." The cited web document contains extensive references to information on XML, including the one quoted below. XML is intended to permit all kinds of structured data such ". . . as spreadsheets, address books, configuration parameters, financial transactions, technical drawings, etc. Programs that produce such data often also store it on disk, for which they can use either a binary format or a text format. The latter allows you, if necessary, to look at the data without the program that produced it. XML is a set of rules, guidelines, conventions, whatever you want to call them, for designing text formats for such data, in a way that produces files that are easy to generate and read (by a computer), that are unambiguous, and that avoid common pitfalls, such as lack of extensibility, lack of support for internationalization/localization, and platform-dependency." (Bos, 1999).

3. Discussion and Conclusions

Meteorological data formats and data representation schemes have proliferated in response to the immense flood of data being generated by satellites, other remote sensors, and numerical weather models. Apparently, efforts to standardize these formats have not kept up with the data flood, consequently those who would make use of all this data are forced to learn how to read, interpret, and display many different data formats. Software packages to manage these tasks have been and continue to be developed, but the day when all data will be transparently available and useable is still in the future.

References

Bos, Bert (1999http://www.w3.org/XML/1999/XML-in-10-points, revised W3C Communications Team (MIT, INRIA, Keio), 2 July, 2001.

Connolly, Dan (1997) "Extensible Markup Language (XML)," http://www.w3.org/XML/, W3C (MIT, INRIA), April, 1997.

Davies, Harvey, (1996) "FAN Utilities document," http://www.unidata.ucar.edu/packages/netcdf/fan_utils.html, UCAR/UNIDATA Program Center, May 20, 1996.

Gudgin, Martin, Marc Hadley, Jean-Jacques Moreau, Henrik Frystyk Nielsen, (2001), http://www.w3.org/TR/2001/WD-soap12-20010709/, July 9, 2001.

Hartnett, Edward (1995). "BUFR, the WMO standard for point data," http://dao.gsfc.nasa.gov/data_stuff/formatPages/BUFR.html, National Aeronautics and Space Agency, December 4, 1995.

Jarvi, A. W. (1996), http://205.156.54.206/oso/oso1/oso12/overview.htm, February 22, 1996.

Jarvi, A. W. (2001a), http://205.156.54.206/noaaport/html/overview.shtml, National Weather Service, November 20, 2001.

Jarvi, A. W. (2001b), "NOAAPORT Data Characteristics," http://205.156.54.206/noaaport/html/data.shtml, National Weather Service, November 20, 2001.

Jones, Andrew S. (2001). "User's Guide for the Data Processing and Error Analysis System (DPEAS)," personal communication. (This exists in html format, but does not seem to be publicly available. Interested parties might try contacting Dr. Jones at Colorado State University.)

Kidwell, Katherine B., (1999) "NOAA Polar Orbiter Data User's Guide, Section 2.0," http://www2.ncdc.noaa.gov/docs/podug/html/c2/sec2-0.htm, as amended December 10, 1999.

King, Joseph H. (2001) "National Space Science Data Center Common Data Format," http://nssdc.gsfc.nasa.gov/cdf/cdf_home.html, NASA Goddard Space Flight Center, February 16, 2001.

Raymond, D. J. (1988). "A C Language-Based Modular System for Analyzing and Displaying Gridded Numerical Data," *Journal of Atmospheric and Oceanic Technology*, 5, 501-511.

Rew, Russ, Glenn Davis, Steve Emmerson, and Harvey Davies (1996), "NetCDF User's Guide," http://www.unidata.ucar.edu/packages/netcdf/guide_11.html#SEC86, Unidata, February, 1996.

Rew, Russ (2001a). "NetCDF," http://www.unidata.ucar.edu/packages/netcdf/, May 31, 2001.

Rew, Russ (2001b). "Frequently Asked Questions About netCDF," http://www.unidata.ucar.edu/packages/netcdf/faq.html#whatisit, June 27, 2001.

Reynolds, J., R. Braden, S. Ginoza, et. al. eds.(2001). "Internet Official Protocol Standards," http://www.globecom.net/ietf/std/std1.html, Network Working Group, Internet Engineering Task Force, August 2001.

Srinivasan, R. (1995). "RFC 1832, XDR: External Data Representation Standard," http://www.faqs.org/rfcs/rfc1832.html, August 1995.

Stackpole, John D. (1993), "GRIB and BUFR," http://ssdoo.gsfc.nasa.gov/nost/formats/bufr.html, National Aeronautics and Space Agency, April, 1993

Thorpe, W. (undated), "A Guide to the WMO Code Form FM 94 BUFR." http://dss.ucar.edu/docs/formats/bufr/bufr.pdf, Fleet Numerical and Oceanography Center, Monterey, CA.

Ullman, Rich (2000). "Earth Observing System (EOS) Data Gateway Glossary of Terms," http://www-v0ims.gsfc.nasa.gov/v0ims/DOCUMENTATION/GLOS-ACR/glossary_of_terms.html, NASA Goddard Space Flight Center EOS Data Gateway Science/Operations Team, February 3, 2000.

American Meteorological Society, WMO Publication Number 306, *Manual on Codes*, Volumes I.1 and I.2 (1995) and Volume II (1998), WMO Publication Number 306 American Meteorological Society, WMO Publications Center, Boston, MA.

European Centre for Medium-Range Weather Forecasts (2001), http://www.badc.rl.ac.uk/data/ecmwf-era/software.html, July 9, 2001.

FMH#4, *Federal Meteorological Handbook No. 4. Radiosonde Code, Second Edition-(*1976) U.S. Department of Commerce, National Oceanic and Atmospheric Administration. Washington, D.C.

FMH#6, *Federal Meteorological Handbook No. 6. Upper wind code (standards and procedures for the coding of upper wind reports)* U.S. Department of Commerce, National Oceanic and Atmospheric Administration. Washington, D.C.

MétéoSum (2001) http://tony.sum.free.fr/metar.htm, MétéoSum, February, 2001.

The National Space Science Data Center, http://nssdc.gsfc.nasa.gov/about/about_nssdc.html

National Center for Supercomputing Applications (1995), ftp://ftp.ncsa.uiuc.edu/HDF/HDF/Documentation/HDF4.1r4/Ref_Manual/January 12, 1995.

National Center for Supercomputing Applications (2001a), http://hdf.ncsa.uiuc.edu/HDF5/doc/Tutor/, June 23, 2001.

National Center for Supercomputing Applications (2001b) Author not listed. http://hdf.ncsa.uiuc.edu/, August 1, 2001.

Naval Research Laboratory (undated), "The BUFR Zone," http://www.nrlmry.navy.mil/bufr/bufr.html

Seaspace (1999) Author not listed. http://www.seaspace.com/service/support/TeraScan_Docs/doc/man7/datasets.html, May 10, 1999.

Seaspace (2000) Author not listed. http://www.seaspace.com/service/support/TeraScan_Docs/doc/intro_frame.html, November 11, 2000.

U.S. Geological Survey web site , http://edcwww.cr.usgs.gov/glis/hyper/guide/avhrr, "Advanced Very High Resolution Radiometer.

University of Wisconsin Space Science and Engineering Center, "McIDAS," http://www.ssec.wisc.edu/software/mcidas.html, May 5, 1997.

Acronyms

URL	Uniform Resource Locator
AVHRR	Advanced very high-resolution radiometer
EOS	Earth Observing System
BUFR	Binary Universal Form
WMO	World Meteorological Organization
GRIB	Gridded binary
CDF	Common Data Format
NASA	National Aeronautics & Space Administration
NSSDC	National Space Science Data Center
EOS	Earth Observing System
HDF	Hierarchical data format
NCSA	National Center for Supercomputing Applications
HRPT	High Resolution Picture Transmission
NOAA	National Oceanic & Atmospheric Administration
POES	Polar Orbiting Environmental Satellites
TIROS-N	Television and Infrared Observation Satellite, N Series
SSEC	Space Science and Engineering Center

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (03-15-2002) February 2003	2. REPORT TYPE Technical	3. DATES COVERED (From - To) July 2001-February 2002

4. TITLE AND SUBTITLE Glossary and Catalog of MeT Data Representation	5a. CONTRACT NUMBER NA
	5b. GRANT NUMBER NA
	5c. PROGRAM ELEMENT NUMBER **611102**
6. AUTHOR(S) Edward M. Measure	5d. PROJECT NUMBER **B53A**
	5e. TASK NUMBER **2HEJ58**
	5f. WORK UNIT NUMBER **WCYIX5**

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Laboratory Information Science and Technology Directorate Battlefield Environment Division (ATTN: AMSRL-CI-EI) White Sands Missile Range, NM 88002-5501	8. PERFORMING ORGANIZATION REPORT NUMBER ARL-TR-2712
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Laboratory 2800 Powder Mill Road Adelphi, MD 20783-1145	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) ARL-TR-2712

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The present short report is intended to catalog the major digital formats in wide use for representing meteorological data. Each data format discussed is briefly defined, and references are given to more detailed discussions of the format. The author intends to follow this report with another that discusses the most important formats used for satellite data in considerably greater detail. No claim is made that the list of formats given is exhaustive, and the author would appreciate readers calling his attention to important formats that have been neglected or overlooked for incorporation into future editions. He can be contacted at emeasure@arl.army.mil

15. SUBJECT TERMS

glossary, meteorological data representations

16. SECURITY CLASSIFICATION OF: Unclassified			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Edward M. Measure
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	SAR	17	19b. TELEPHONE NUMBER (Include area code) (505) 678-3307